

The effects of ICAP learning activities on learning outcomes

Christina Wekerle, Michael Sailer, Ingo Kollar, Tim Fütterer

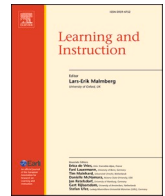
Angaben zur Veröffentlichung / Publication details:

Wekerle, Christina, Michael Sailer, Ingo Kollar, and Tim Fütterer. 2027. "The effects of ICAP learning activities on learning outcomes." *Learning and Instruction* 107: 102482.
<https://doi.org/10.1016/j.learninstruc.2026.102482>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0





The effects of ICAP learning activities on learning outcomes

Christina Wekerle^{a,*}, Michael Sailer^a, Ingo Kollar^a, Tim Fütterer^b

^a University of Augsburg, Universitätsstraße 10, Augsburg, 86159, Germany

^b University of Tübingen, Hector-Institut für Empirische Bildungsforschung, Europastraße 6, Tübingen, 72072, Germany

ARTICLE INFO

Keywords:

ICAP
Learning activity
Knowledge
Motivation
Cognitive processes

ABSTRACT

Background: The Interactive-Constructive-Active-Passive (ICAP) framework makes assumptions about the relationships between observable types of learning activities (passive, active, constructive, interactive), cognitive processes, and learning outcomes. Yet, rigorous empirical studies are still lacking establishing the validity of the ICAP framework.

Aims: This study investigates the effects of student engagement in passive, active, constructive and interactive learning activities on their surface and deep knowledge gain, and their motivation. Additionally, the study aims to determine the associations between students' cognitive processes, their learning activities, as well as surface and deep knowledge gain.

Sample: Participants were 249 pre-service teachers.

Methods: After reading a text on multimedia learning, participants were asked to reread the text (passive condition), to underline central information in the text (active condition), to work on three open-ended tasks addressing the text (constructive condition), or to work on the same tasks in pairs (interactive condition).

Results: No statistically significant effects of condition on surface or deep knowledge could be detected. Regarding motivation, participants in the interactive condition reported statistically significantly higher levels of perceived relatedness compared to all other conditions. Furthermore, only participants in the interactive condition reported statistically significantly higher levels of interactive cognitive processes compared to those in all other conditions. Constructive cognitive processes were statistically significantly negatively associated with deep knowledge gain.

Conclusion: The findings suggest that the empirical foundation supporting the ICAP framework may not yet be sufficiently robust. Yet, more research is needed that applies a more rigorous methodology exploring the ICAP framework.

1. Introduction

For any pedagogical practice, an understanding of central learning mechanisms is essential to teachers. In addition to seminal theories, such as Self-Determination Theory (Ryan & Deci, 1985) and Cognitive Load Theory (Sweller, 1988), the scientific community has also started to shape its understanding of learning and its effects over the last decade through the Interactive-Constructive-Active-Passive (ICAP) framework. According to Google Scholar, the two most important articles were cited over 2000 (Chi, 2009) and 4000 times (Chi & Wylie, 2014).

The ICAP framework makes assumptions about the relationships between various observable types of learning activities (passive, active, constructive, interactive), cognitive processes, and learning outcomes.

At the heart of the framework is the ICAP hypothesis, which states that these types of learning activities should be conducive to learners' learning success in the following order: interactive > constructive > active > passive. To directly test this hypothesis based on reanalyses of previous studies, Menekse et al. (2013) conducted an initial between-subject laboratory study, which confirmed the differential effectiveness of the types of learning activities for learning outcomes in the assumed order (Menekse et al., 2013). Whereas this study represents a first critical step in validating the ICAP hypothesis, various methodological issues limit the explanatory power of the results (Thurn et al., 2023).

Other between-subject experimental studies have followed since then, making important contributions to advancing our understanding

* Corresponding author. University of Augsburg, Universitätsstraße 10, Augsburg, 86159, Germany.

E-mail addresses: christina.wekerle@uni-a.de (C. Wekerle), michael.sailer@uni-a.de (M. Sailer), ingo.kollar@uni-a.de (I. Kollar), tim.fuetterer@uni-tuebingen.de (T. Fütterer).

<https://doi.org/10.1016/j.learninstruc.2026.102482>

Received 18 November 2025; Received in revised form 12 August 2026; Accepted 25 September 2026

Available online 9 October 2026

0959-4752/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Open access funding enabled and organized by Projekt DEAL.

of the ICAP framework by elucidating how interactive, constructive, active, and passive learning activities affect cognitive processes and outcomes (e.g., Liu et al., 2024; Xie et al., 2024). However, most of these studies (1) only included some of the four types of learning activities (e.g., Petersen et al., 2023), (2) did not control for time-on-task (e.g., Peterson et al., 2023), (3) did not differ between superficial and deep knowledge gain as outcome variables despite assumptions in the ICAP theory (e.g., Liu et al., 2024), (4) did not consider relevant learning outcomes (such as students' motivation) besides knowledge gain (e.g., Lam & Muldner, 2017), and (5) results were only partly in line with the ICAP hypothesis and consequently cast doubts on its validity.

Therefore, in the present pre-registered study (<https://doi.org/10.17605/OSF.IO/3MVCA>), we investigated the effects of the four different types of learning activities as they were implemented in the initial ICAP-based study (Menekse et al., 2013) on students' shallow and deep knowledge gain and motivation, while taking important limitations of the original study into account.

2. The ICAP framework

Rooted in discussions on advancing so-called active learning in teaching, Chi and Wylie (2009; 2014) developed the Interactive-Constructive-Active-Passive (ICAP) framework to address the mostly unspecific ideas presented in the literature on active learning. The intention of the framework is to provide teachers with an observable, evidence-based categorization of four types of learning activities that differ in their importance for cognitive learning outcomes, thereby allowing to choose appropriate instructional approaches, determine the quality of learning products, and therefore increase the chances of learners' deep engagement with instructional material (Chi & Wylie, 2014). The underlying assumption is that certain learning activities increase the likelihood of certain cognitive processes occurring, which in turn should lead to qualitatively different cognitive learning outcomes. In the following, the four observable types of learning activities are presented along with examples, presumed cognitive processes that accompany them, and probable cognitive learning outcomes.

Passive learning activities involve learners focusing on the material to be learnt and receiving information without any other observable behaviour. Examples of passive learning activities are reading a textbook entry, listening to a podcast, or watching a video. This type of activity should allow for *storing* information in an isolated manner. Consequently, learners might only *recall* information when presented in a context identical to the acquisition context.

Compared to passive learning activities, *active* learning activities involve physical interaction with the learning material, e.g., by highlighting information in a text, taking verbatim notes of a podcast, or by stopping or forwarding a video. This form of engagement with the learning material is supposed to encourage learners to *integrate* information with their prior knowledge. Therefore, they should be able to *apply* their knowledge to non-identical, but similar contexts.

Constructive learning activities are characterized by learners creating learning products that go beyond the initially presented instructional material. These learning products might consist of crafting a summary in one's own words, transferring questions or answers addressing the topic of a podcast, or constructing a visualization of a video's topic. The framework assumes that the constructive type enables learners to *infer* knowledge from their prior knowledge and linked new information by means of revising and restructuring knowledge. As a result of learning, learners might *transfer* their knowledge to new contexts.

In contrast to the other three types, *interactive* learning activities involve two or more learning partners interacting with each other by referring to each other's utterances and continuously taking turns. Interactive learning activities might take place when learners discuss the pros and cons of the process described in a textbook entry in a debate, jointly creating a concept for a follow-up podcast episode, or giving each other feedback and explaining parts of a video to each other. At the

cognitive level, these activities may enable *co-inferring* processes due to additional contributions or information of the learning partners next to learners' own elaboration of the learning material. In terms of cognitive outcomes, learners might *co-create* new, creative ideas and solutions.

Consequently, the ICAP hypothesis reads as follows: Interactive learning activities should be more conducive to deeper learning than constructive learning activities. Constructive learning activities, in turn, should result in deeper learning than active learning activities. Finally, active learning activities are assumed to produce deeper learning than passive learning activities (Chi & Boucher, 2023).

However, certain preconditions have to be considered when applying the ICAP hypothesis to (research) practice:

As described above, the ICAP hypothesis rests on the assumption that engagement in different observable learning activities is related to different levels of cognitive processing. Yet, Chi and Wylie (2014) do not assume a mechanistic relation between observable activities and cognitive processing, but rather a probabilistic one. In other words, at least in some situations learners may engage in deep elaborative processes while listening to an interesting lecture, just as they may engage in shallow cognitive processes during collaborative learning. Nevertheless, according to Chi and Wylie (2014), the probability of deep cognitive processing increases from passive to active to constructive to interactive activities.

Also, Chi et al. (2018) consider the difference between interactive/constructive, and active/passive learning activities to be larger than between interactive and constructive on the one hand, and active and passive learning activities on the other hand. This distinction is based on whether the respective activities are assumed to involve predominantly shallow or deep processing.

The idea of distinguishing between shallow and deep cognitive processing (ultimately resulting in shallow and deep knowledge as a product of such processing) has long been introduced in traditional information-processing theories such as Craik & Lockhart (1972), Craik & Tulving (1975), and Marton & Saljo (1976). However, while these theories primarily describe differences in processing depth, they provide limited guidance for identifying observable learning activities that are likely to elicit such processing and for translating these distinctions into measurement approaches. The ICAP framework provides a useful way to establish such this link by relating observable learning activities to different forms of cognitive engagement. Accordingly, and in line with the probabilistic, rather than mechanistic relation between observable learning activities and cognitive processes, we tentatively propose to view the cognitive processes that Chi and Wylie associate with passive (i.e., isolated storing) and active (i.e., integrating information with prior knowledge) learning activities involving rote processing as "shallow" (Chi et al., 2018; Greene, 2015; Bauer et al., 2025). By contrast, constructive and interactive learning activities are associated with elaborative processes, namely inferring and co-inferring, and can therefore be conceptualized as "deep processing" (Chi et al., 2018; Bauer et al., 2025; Wekerle et al., 2024).

In turn, we assume that engagement in shallow processing is most likely to result in shallow knowledge, whereas engagement in deep processes is most likely to result in deep knowledge. Following common distinctions between knowledge retention and knowledge transfer (Mayer, 2009), *shallow knowledge* refers to the retention of factual information, definitions, and concepts. *Deep knowledge*, by contrast, encompasses the ability to apply knowledge to novel situations. This perspective is particularly valuable for the present study because our measurement approach requires observable indicators for underlying cognitive processing.

3. Empirical findings and shortcomings of the ICAP framework

To substantiate the ICAP hypothesis, Chi (2009) and Chi and Wylie (2014) conducted reanalyses of previous laboratory and field studies by categorizing the experimental conditions according to the ICAP types

of activities (e.g., constructive vs. active). The comparisons were mainly based on studies relating to note-taking, concept mapping, and self-explanations. The approximately 40 listed studies can be considered a solid empirical basis in favour of the ICAP hypothesis. However, it should be noted that the reanalysis was not intended as a systematic review. Therefore, it lacks information on the search methodology, relies on interpretations of experimental conditions not designed based on the ICAP framework, and exclusively presents ICAP hypothesis-compliant studies. For this reason, it seems necessary to also evaluate experimental studies that have been performed to directly test the ICAP hypothesis.

The initial and to date only experimental study known to the authors that tested the effects of all four types of learning activities in a between-subjects design was performed by Menekse et al. (2013). In this study, after a brief introduction to the content and a knowledge pretest, engineering students were assigned to four different conditions. In the passive condition, students were asked to read a text aloud. In the active condition, students were prompted to underline the most important sentences in this text. In the constructive condition, students were to answer questions on a worksheet using diagrams and illustrations based on the same text as in the other two conditions. In the interactive condition, dyads of students worked on the same worksheet as in the constructive condition, but were instructed to come to a consensus in their group before writing their answers down. After the intervention, students' knowledge was tested again with questions from the pre-test and additional questions. The results demonstrated a statistically significant increase of knowledge gain from type of activity to type of activity (d (I vs. C) = 0.64, d (C vs. A) = 0.54, d (A vs. P) = 0.65), fully in line with the ICAP hypothesis.

Besides the initial experimental study, which investigated all four types of learning activities, most of the subsequent experimental between-subject studies typically compared only two types of learning activities. Here, the pattern of results is inconsistent. Whereas some studies support the proposed ICAP hierarchy (e.g., C > P: Liu et al., 2024; A > P: LaDage et al., 2018), others report statistically non-significant differences in learning outcomes (e.g., I vs. C: Petersen et al., 2023; C vs. A: Lam & Muldner, 2017). In some cases, even comparisons between learning activities positioned farther apart in the hierarchy yielded statistically non-significant effects (e.g., I vs. A: Luo et al., 2016).

However, several limitations of these studies warrant further consideration:

- (1) A key limitation, particularly in the case the Menekse et al. (2013), lies in the variation of learning materials across conditions. Whereas students in the passive and active conditions received an eight-page text, those in the constructive and interactive groups did not work with the text, but with visual summaries of the same content, which may have provided them with an informational advantage.
- (2) Whereas variations in time-on-task can affect learning outcomes, this factor has not been controlled for in the reviewed studies – except for Lam and Muldner (2017). For example, although all students were allocated 25–30 min in the study of Menekse et al. (2013), those in the passive condition likely spent less time engaging with the material than peers with more demanding tasks.
- (3) Additionally, despite being suggested to do so, most knowledge tests implemented in the mentioned studies (except for Lam & Muldner, 2017; Petersen et al., 2023) did not differentiate between shallow and deep knowledge.
- (4) Furthermore, students' knowledge-change or cognitive processes are considered a necessary factor mediating the relationship between observable learning activities and learning outcomes in the ICAP framework. Therefore, in verifying the ICAP hypothesis and

to better understand conflicting results, they should be additionally assessed. Yet, they have rarely been investigated.

- (5) Finally, the ICAP framework only focuses on cognitive learning outcomes. In contrast, empirically based teaching and learning frameworks point to the importance of considering cognitive as well as affective-motivational learning outcomes to determine the effectiveness of instructional approaches (e.g., Kunter et al., 2013). Consequently, the first ICAP-based studies assessed the association between learning activities and affective-motivational outcomes. These studies identified some positive effects of learning activities at higher hierarchical levels compared to activities at lower higher hierarchical levels on affective-motivational outcomes (Conley et al., 2022: A > P on engagement, C & I > P on frustration; Wekerle et al., 2024: C > P on joy; Xie et al., 2024: C > P, C > A on motivation). Once again, between-subject experimental designs have yet to be implemented.

These implementation inconsistencies highlight the need for an experimental between-subjects study that (1) compares all four types of learning activities, (2) uses the same materials, (3) controls for task duration, (4) uses shallow and deep knowledge tests, (5) considers students' reported cognitive processes along with their associations to learning activities and learning outcomes, and (6) assesses students' affective-motivational outcomes along with cognitive outcomes.

4. Research questions and hypothesis

Based on the ICAP framework and the findings presented above, we investigated the effects of student engagement in the four types of learning activities on their cognitive and motivational learning outcomes. Additionally, we aimed to investigate to which degree self-reported cognitive processes are associated with students' learning activities and their learning outcomes. In the following, we present five research questions (RQ), which we preregistered on OSF (<https://doi.org/10.17605/OSF.IO/3MVCA>). All deviations from the preregistration can be found in the OSF project (<https://doi.org/10.17605/OSF.IO/KJU36>). For the first RQ, we present a hypothesis. The other four RQs are of an exploratory nature.

RQ1: Do interactive learning activities have a stronger positive effect than constructive learning activities, do constructive have a stronger positive effect than active learning activities, and do active have a stronger positive effect than passive learning activities on learners' deep knowledge gain?

H: Based on the ICAP hypothesis of Chi and Wylie (2014) as well as the results of the laboratory study of Menekse et al. (2013), we expected interactive learning activities to have a stronger statistically positive effect than constructive learning activities, which we expected to have a stronger positive effect than active learning activities, which we expected to have a stronger positive effect than passive learning activities on learners' deep knowledge gain from T1 to T2 (ICAP-hypothesis).

RQ2: Do interactive learning activities have a stronger positive effect than constructive learning activities, do constructive have a stronger positive effect than active learning activities and do active have a stronger positive effect than passive learning activities on learners' shallow knowledge gain?

RQ3: Do interactive learning activities have a stronger positive effect than constructive learning activities, do constructive learning activities have a stronger positive effect than active learning activities, and do active learning activities have a stronger positive effect than passive learning activities on students' reported motivation?

RQ4: To which degree do participants in one condition (e.g., passive) report having the corresponding type of cognitive process (e.g., passive)

performed to a statistically significantly higher degree than participants in the other three learning activity conditions?
RQ5: To which degree are participants' reported types of cognitive processes predictive of participants' knowledge gain?

5. Method

5.1. Sample

Our sample consisted of $N = 247$ pre-service teachers at a medium-sized university in the south of Germany. The sample was collected in two waves (February 2023 and February 2025). The students participated in the study in the context of a mandatory introductory Psychology lecture. Therefore, participation in the learning task was mandatory. However, the provision of data was voluntary. All participants gave informed consent prior to participation, and the study was conducted in accordance with the German Psychological Society's (DGPS) ethical guidelines. 78% of students self-identified as female. The mean age of the participants was $M = 20.74$ years ($SD = 3.22$, $Mdn = 20.00$; range = 18–42). Participants were predominantly in their first semester of study (80%), with a smaller number from later semesters.

Participants reported a wide range of subject specializations, reflecting the heterogeneity of teacher education in Germany. Teaching subjects included German, English, mathematics, geography, and combinations thereof. Likewise, participants' didactic specializations spanned subjects such as mathematics, German, English, arts, religion, and social studies.

5.2. Design

In accordance with Menekse et al. (2013), we investigated the factor type of learning activity with four conditions in a between-subject design. Based on their preference, participants applied for different study time slots via an online form and were then assigned to one time slot. Each time slot was randomly assigned to one of the following four conditions: passive ($n = 52$), active ($n = 49$), constructive ($n = 46$), and interactive ($n = 100$).

We conducted preliminary analyses to examine baseline equivalence across conditions. An ANOVA indicated no statistically significant differences regarding age between conditions, $F(3, 238) = 0.52$, $p = .670$, $\eta^2 = 0.01$. Similarly, a chi-square test showed no statistically significant differences in gender distribution across conditions, $\chi^2(3) = 4.56$, $p = .207$. This suggests that randomization was successful in generating comparable groups at baseline.

A priori power analyses using G*Power (version 3.1.9.4) indicated that a total sample size of 225 participants would be sufficient to detect effects of medium size ($f = 0.25$) with 80% power and an alpha level of 0.05 in a one-way between-subjects ANOVA. To account for potential exclusions and to increase statistical power, we aimed to collect data from as many students as possible.

5.3. Procedure

The study was conducted in small groups under supervised and standardized conditions in one of the university's computer labs. At the beginning of the session, the experimenter provided standardized instructions regarding the working procedure, technical aspects of the online environment, and the overall study sequence. Before starting the pretest, participants additionally received written information within the online environment about the study's purpose and procedure, were informed about data protection, and provided informed consent for the processing of their personal data. After an online-based pretest with 18 multiple choice questions on multimedia learning measuring shallow and deep knowledge, all participants were instructed to first read a four-pages long, paper-based text on multimedia learning (adapted from Leutner et al., 2014). The text focused on the Cognitive Theory of

Multimedia Learning (Mayer, 2009), Cognitive Load Theory (Sweller et al., 1998), and related principles for designing multimedia learning environments. Participants from all conditions were prompted to read the text carefully within 10 min, to work alone and to not underline or make any notes in the text. Next, the intervention was scheduled to last for 15 min (see manipulation of the independent variable). Immediately after the intervention, all participants completed an online-based knowledge posttest to measure shallow and deep knowledge featuring the same questions as in the pretest. Item order was randomized for both test administrations. Subsequently, students reported on their motivation, effort, elaboration strategies, and types of cognitive processes. Finally, participants provided their demographic information. The procedure including approximate durations is illustrated in Table 1.

All materials were pre-tested with three students who were employed as student research assistants. This group was chosen because they combine the perspective of the target population (students) with sufficient familiarity with academic materials. Students worked through all materials under conditions comparable to the planned implementation, and provided feedback on clarity, perceived difficulty, usability, and overall length. Based on this feedback, minor revisions were made.

5.4. Manipulation of the independent variable

In the passive condition, participants were asked to read the paper-based text a second time, using the same instructions as before. In the active condition, participants were asked to underline the central information in the paper-based text, to work alone and not to make any notes in the text. Three differently coloured markers were provided. In the constructive condition, participants were asked to work on three open-ended tasks addressing the text on a paper-based worksheet. First, students should explain the most important findings on multimedia learning in their own words. Second, they were to write down two examples illustrating the types of cognitive load and design principles. Third, three conclusions based on findings for own teaching had to be elaborated. In the interactive condition, participants were asked to work on the same three open-ended tasks as those in the constructive condition, but in pairs. They were asked to establish consensus before writing down their answers, and their conversations were recorded.

All participants were also required to indicate at the end of the intervention whether they had fully completed the task. Only if they confirmed, they were forwarded to the posttest. To ensure that the length of the intervention was comparable between conditions, the test administrator provided time-based completion instructions in addition to the time information in the online environment. All prompts, tasks, and questions of the intervention can be found in the preregistered OSF

Table 1
Overview of study procedure and duration of each phase.

Phase	Description	Duration
Pretest	Multiple-choice questions on multimedia learning	ca. 13 min
Initial Reading	Reading of a four-page text on multimedia learning	ca. 10 min
Intervention (by condition)	- Passive: reread text - Active: underline central information - Constructive: work on 3 open-ended tasks - Interactive: work on 3 open-ended tasks in pairs	ca. 15 min
Posttest	Multiple-choice questions on multimedia learning Motivation Effort Elaboration strategies Types of cognitive processes	ca. 13 min
Demographics	Collection of demographic information	ca. 4 min

project.

5.5. Measures

Deep Knowledge. We measured students' deep knowledge with nine self-developed multiple-choice questions on multimedia learning, with four items each, resulting in a total of 36 items. The questions were developed based on the central theoretical elements of the relevant frameworks presented in the learning material on multimedia learning, ensuring that all key aspects of the theories were represented. For both shallow and deep knowledge, an equal number of questions were constructed for each targeted theoretical element. The validity of the knowledge type categorization was implemented in a consensus process by two content experts. Items focused on students' application of multimedia learning theories and principles derived from [Leutner et al. \(2014\)](#) to different practical scenarios. Example question: "In mathematics class, multiplication of fractions is taught. In a subsequent test, Max performs poorly. According to Cognitive Load Theory, what are possible causes for his poor performance? (1) Discussing solution strategies with others reduced his germane cognitive load. (2) During the lesson in which the topic was covered, Max was not sufficiently able to invest free working memory capacity in germane cognitive load. (3) Intrinsic cognitive load was too low for Max during the test session. (4) Extraneous cognitive load was too high for Max during the lesson in which the topic was covered." Pre- and posttest items were identical to ensure direct comparability of knowledge scores across both measurement points. Using the same items allows for an assessment of knowledge gain attributable to the intervention, rather than differences in item difficulty or content. For a correct answer, participants received one point; for an incorrect response, they received zero points. Items with an item difficulty below 0.20 or above 0.80 were excluded, except for 20 predefined items considered theoretically important, which were retained regardless of item difficulty (see the results in our OSF project). Items outside the predefined difficulty range were retained if they captured theoretically central aspects of the construct that were deemed essential for content validity, ensuring that the composite score reflects the full conceptual breadth of students' deep knowledge rather than being restricted solely by psychometric criteria. Of the 32 items we included, the mean score served as the measure of deep knowledge (i.e., a maximum of 32 points). Whereas Omega Hierarchical was $\omega_H = 0.55$ at T1 and $\omega_H = 0.60$ at T2, indicating that 55% (60%) of the variance in total scores could be attributed to a general factor, Guttman's λ_6 indicated a higher internal consistency (T1: $\lambda_6 = 0.76$; T2: $\lambda_6 = 0.92$), suggesting shared variance across items despite potential multidimensionality. To evaluate internal consistency, we additionally computed McDonald's Omega Total, which was $\omega_{\text{total}} = 0.55$ at T1 and $\omega_{\text{total}} = 0.84$ at T2. This suggests that approximately 55% (84%) of the variance in observed total scores is attributable to true score variance, including both general and specific latent factors. The differences in reliability estimates between T1 and T2 likely reflect changes in the underlying measurement structure, with lower coefficients at T1 indicating greater heterogeneity in students' initial knowledge representations, and higher consistency at T2 suggesting increased coherence of the construct following the intervention, although the pattern of coefficients overall points to some degree of multidimensionality. The items can be accessed via the OSF project.

Shallow Knowledge. We measured shallow knowledge with nine self-developed multiple-choice questions on multimedia learning, with four items each, resulting in a total of 36 items. The items targeted factual recall of information, definitions, and concepts. Example question: "Which statement(s) are compatible with findings on multimedia learning? (1) The processing capacity of human working memory is limited. (2) Subjective cognitive load should not exceed working memory capacity (objective cognitive load). (3) Free working memory capacity should, if possible, be invested in learning-relevant cognitive load. (4) Content-related cognitive load can be distributed between

working memory and long-term memory." The variable was measured at pre- and posttest. Scoring and exclusion criteria were identical to those for deep knowledge. Of the 34 items we included, the mean score served as the measure of shallow knowledge (i.e., a maximum of 34 points). Whereas Omega Hierarchical was $\omega_H = 0.58$ at T1 and $\omega_H = 0.71$ at T2, indicating that 58% (71%) of the variance in total scores could be attributed to a general factor, Guttman's λ_6 indicated a higher internal consistency (T1: $\lambda_6 = 0.84$; T2: $\lambda_6 = 0.85$), suggesting shared variance across items despite potential multidimensionality. To evaluate internal consistency, we additionally computed McDonald's Omega Total, which was $\omega_{\text{total}} = 0.61$ at T1 and $\omega_{\text{total}} = 0.78$ at T2. This suggests that approximately 61% (78%) of the variance in observed total scores is attributable to true score variance, including both general and specific latent factors. The items can be accessed via the OSF project.

Motivation. The measurement of participants' motivation was based on self-determination theory ([Deci & Ryan, 1985](#)). We measured participants' motivation in the posttest with five Likert-scales on a scale from 1 (*do not agree at all*) to 5 (*totally agree*). Three (*interest*, *perceived competence*, and *perceived choice*) out of four subscales from [Wilde et al. \(2009\)](#) were used to measure motivation. The subscale *pressure* was not included, as it does not align with the theoretical focus on self-determined motivation. However, the scales *social belonging* and *perceived task importance* were included to comprehensively capture motivational constructs relevant to self-determination theory ([Sailer et al., 2017](#)). The original items refer to experiences in an exhibition; for the present study, this wording was adapted to "learning content". The scales were the following (1) *interest*: three items (e.g., "I found the learning content very interesting."; [Wilde et al., 2009](#)); with an internal consistency of $\alpha = .81$ (2) *perceived competence*: three items (e.g., "I am satisfied with my performance in dealing with the learning content."; [Wilde et al., 2009](#)); with an internal consistency of $\alpha = .80$ (3) *perceived choice*: three items (e.g., "When dealing with the learning content, I was able to choose how I did it"; [Wilde et al., 2009](#)); with an internal consistency of $\alpha = .83$ (4) *social belonging*: three items (e.g., "I felt like part of a team."; [Sailer et al., 2017](#)); with an internal consistency of $\alpha = .81$ (5) *perceived task importance*: three items (e.g., "It was worthwhile doing this activity."; [Sailer et al., 2017](#)); with an internal consistency of $\alpha = .90$.

Types of Cognitive Processes. We measured participants' types of cognitive processes adapted from [Stegmann et al. \(2024\)](#) with four Likert-scales on a rating scale from 1 (*does not apply*) to 5 (*totally applies*): (1) *passive cognitive processes (storing)*: three items (e.g., "I tried to memorize the learning content exactly as it was presented in the theoretical text.") with an internal consistency of $\alpha = .73$; (2) *active cognitive processes (integrating)*: three items (e.g., "I thought about how the learning content related to what I already knew about the topic.") with an internal consistency of $\alpha = .63$; (3) *constructive cognitive processes (inferring)*: three items (e.g., "I reorganized the learning content in a way that made sense to me.") with an internal consistency of $\alpha = .63$; (4) *interactive cognitive processes (co-inferring)*: four items (e.g., "I reflected on whether the other person's contributions to the learning content were correct.") with an internal consistency of $\alpha = .84$. Although Cronbach's alpha was $\alpha = .63$ for the active and constructive cognitive process scales, the mean inter-item correlations were 0.37 and 0.36, respectively, falling within the recommended range of 0.20–0.40 for short scales and indicating adequate internal consistency despite the limited number of items ([Piedmont, 2023](#)). We assume that the partially low alpha values are because the scales test a selection of different cognitive processes rather than individual, specific cognitive processes. Therefore, it is questionable whether alpha can be used as an adequate indicator of reliability ([Stadler et al., 2021](#)). The items can be accessed via the OSF project.

Time-on-Task. To assess the duration of students' engagement with the intervention, we calculated a time-on-task variable based on system-generated timestamps. Specifically, we subtracted the recorded start time from the end time for each participant. The resulting value reflects the total amount of time (in seconds) that students actively spent

interacting with the intervention materials.

Whenever we used scales, we utilized manifest scale scores (i.e., arithmetic mean).

5.6. Analyses

We conducted all analyses using R (v4.5.0, R Core Team, 2023) and utilized RStudio (Posit team, 2023).

For statistical analyses, the condition variable was treated as a categorical factor and effect coded. This approach was implemented in R via *contr.sum(4)*, after defining condition as a factor with levels 1–4. In effect coding, the reference group is represented by the negative sum of the coded values for the other groups, allowing each group's effect to be interpreted relative to the overall grand mean rather than to a single reference category.

To address RQ1 and RQ2 concerning the effects of the experimental conditions on knowledge gains, we performed ANOVAs and ANCOVAs using gain scores for deep and shallow knowledge as dependent variables using the *anova* function in R. Next, we performed post-hoc contrasts using the function *emmeans* from the R package *emmeans* (v1.11.1, Lenth, 2024) for pairwise comparisons of conditions.

RQ3 and RQ4 were examined using MANOVAs to test the effects of the ICAP condition on self-reported motivation (i.e., interest, competence, choice, social belonging, and task importance) and reported types of cognitive processes (i.e., passive, active, constructive, and interactive). Univariate follow-up tests were conducted when multivariate effects were statistically significant.

RQ5 investigated the predictive value of the different reported types of cognitive processes (passive, active, constructive, interactive) on both deep and shallow knowledge gains using multiple linear regression analyses. Standardized beta coefficients were computed, and multicollinearity was examined using the variance inflation factor.

Handling Missing Data. Across all indicators at both measurement points, 0.9% of the data were missing (based on 44,909 entries in total). When considering only the scale scores, the proportion of missing data was 1.0%. If individual scale items showed missing values, we calculated the scale mean based on the available values (i.e., *na.rm = TRUE* in R). Furthermore, we did not take any measures to estimate or impute missing values in the analyses.

Robustness Checks. To test the robustness of our findings regarding RQ1 and RQ2, we conducted two robustness checks. As a first robustness check, students' *deep knowledge* and *shallow knowledge* scores at T1 and *time-on-task* were included as covariates in our models (i.e., ANCOVA). More precisely, we tested students' knowledge assessed at T1 for baseline inequivalence according to the What Works Clearinghouse Standards (i.e., $0.00 \leq |\text{ES at baseline}| \leq 0.05$ satisfies the baseline equivalence standard, $0.05 < |\text{ES at baseline}| \leq 0.25$ requires statistical adjustment to satisfy the baseline equivalence standard, $|\text{ES at baseline}| > 0.25$ does not satisfy the baseline equivalence standard; What Works Clearinghouse, 2022) by performing one-way ANOVAs. Additionally, we examined whether students' time-on-task differed between conditions.

As a second robustness check, we performed linear mixed-effects models (LMM) with random intercepts for participants to account for repeated measures over time. These LMMs included interaction terms between time and condition and were estimated using the *lmer* function from the R package *lmerTest* (v3.1.3, Kuznetsova et al., 2017).

We controlled for false discovery rates in all our models by applying the Bonferroni-Holm (1979) correction per model. Effect sizes were reported using eta squared (η^2) for ANOVA-based models and Cohen's *d* for pairwise comparisons. The significance level was set to $\alpha = .05$ in all our models. This study, including the approaches to analyse the data, was preregistered at OSF (<https://doi.org/10.17605/OSF.IO/3MVCA>). The full reproducible code (i.e., we used RMarkdown and the *renv* package) and data are online in our OSF project (<https://doi.org/10.17605/OSF.IO/KJU36>).

6. Results

6.1. Preliminary results

To provide an overview of all central independent, dependent, and control variables, we report descriptive statistics (e.g., *M*, *SD*) by condition and the results of baseline equivalence (Table 2). The correlations of among the employed scales per condition can be found in the appendix.

Descriptively, mean deep knowledge scores increased from T1 to T2 across all conditions (see Fig. 1). Similarly, shallow knowledge (see Fig. 2) showed overall positive tendencies—except for the passive condition.

No statistically reliable differences between conditions emerged for either deep nor shallow knowledge ($ps = 0.475$ and 0.896 , respectively). Nevertheless, the observed effect sizes—although small—were in a direction consistent with slightly higher values for the active and interactive conditions compared to the passive and constructive conditions for deep knowledge ($d = 0.20$), and for the active, constructive, and interactive conditions compared to the passive condition for shallow knowledge ($d = 0.10$).

6.2. Effect of types of learning activities on students' deep knowledge gain

To examine the ICAP hypothesis (RQ1), an ANOVA was conducted with learning condition (i.e., four treatment groups: passive, active, constructive, and interactive) as independent variable and deep knowledge gain as dependent variable. The overall ANOVA was not statistically significant, $F(3, 243) = 1.58$, $p = .194$, indicating no statistically significant differences in deep knowledge gains between the four ICAP conditions. Pairwise comparisons using Bonferroni-Holm correction and Cohen's *d* revealed small, statistically non-significant differences between conditions ($ps > 0.05$, Cohen's *ds* ranging from 0.05 to 0.34), except for the difference between the passive and the interactive condition ($d = 0.34$, 95% CI $[-0.68, -0.01]$, $p = .046$). The interactive condition showed a more substantial gain than the passive condition.

This statistically significant difference between the passive and the interactive condition was not robust in our robustness checks. That is, the ANOVA revealed no statistically reliable differences in deep knowledge gains across ICAP conditions, with only small, mostly non-significant pairwise effects and a non-robust advantage of the interactive over the passive condition. Thus, the ICAP hypothesis was not supported.

6.3. Effect of types of learning activities on students' shallow knowledge gain

To explore the ICAP hypothesis regarding students' acquisition of shallow knowledge (RQ2), we conducted an ANOVA with learning condition (i.e., four treatment groups: passive, active, constructive, interactive) as independent variable and shallow knowledge gain as dependent variable. The overall ANOVA was not statistically significant, $F(3, 243) = 0.81$, $p = .489$, indicating no statistically significant differences in shallow knowledge gains across the four ICAP conditions. Pairwise comparisons using Bonferroni-Holm correction and Cohen's *d* indicated small, statistically non-significant differences between conditions ($ps > 0.05$, Cohen's *ds* ranging from 0.04 to 0.30). For instance, the contrast between the active and constructive condition yielded $d = 0.20$, 95% CI $[-0.20, 0.61]$.

The pattern of results did not change in the robustness checks. That is, the ANOVA revealed no statistically reliable differences in shallow knowledge gains across ICAP conditions, with only small, non-significant pairwise effects and no change in results across robustness checks.

Table 2

Descriptive statistics by condition of all scales used in this study and baseline equivalence.

Variable	Condition	N	M	SD	Median	Min	Max	SE	Baseline Equivalence
Deep knowledge T1 ^a	P	52	0.44	0.10	0.45	0.25	0.70	0.01	$p = .475$
Deep knowledge T1 ^a	A	49	0.47	0.10	0.45	0.25	0.70	0.01	$d = 0.20$
Deep knowledge T1 ^a	C	46	0.44	0.08	0.45	0.25	0.65	0.01	
Deep knowledge T1 ^a	I	100	0.45	0.10	0.45	0.25	0.80	0.01	
Deep knowledge T2 ^a	P	52	0.48	0.59	0.59	-2.41	0.78	0.08	-
Deep knowledge T2 ^a	A	49	0.56	0.44	0.62	-2.41	0.75	0.06	
Deep knowledge T2 ^a	C	46	0.52	0.45	0.58	-2.41	0.84	0.07	
Deep knowledge T2 ^a	I	100	0.62	0.10	0.62	0.34	0.88	0.01	
Shallow knowledge T1 ^b	P	52	0.59	0.10	0.60	0.29	0.74	0.01	$p = .896$
Shallow knowledge T1 ^b	A	49	0.60	0.08	0.60	0.43	0.77	0.01	$d = 0.10$
Shallow knowledge T1 ^b	C	46	0.60	0.08	0.60	0.46	0.89	0.01	
Shallow knowledge T1 ^b	I	100	0.60	0.10	0.60	0.34	0.80	0.01	
Shallow knowledge T2 ^b	P	52	0.59	0.15	0.62	0.00	0.76	0.02	-
Shallow knowledge T2 ^b	A	49	0.64	0.12	0.65	0.00	0.79	0.02	
Shallow knowledge T2 ^b	C	46	0.61	0.12	0.62	0.00	0.79	0.02	
Shallow knowledge T2 ^b	I	100	0.61	0.08	0.62	0.41	0.79	0.01	
Time-on-task ^c	P	40	718.80	168.65	724.50	182.00	1027.00	26.67	-
Time-on-task ^c	A	48	916.54	107.44	943.50	506.00	1114.00	15.51	
Time-on-task ^c	C	44	885.02	104.12	899.50	607.00	1111.00	15.70	
Time-on-task ^c	I	81	957.10	213.50	1030.00	119.00	1184.00	23.72	
Interest ^d	P	51	3.30	0.97	3.33	1.33	5.00	0.14	-
Interest ^d	A	48	3.12	0.83	3.00	1.33	5.00	0.12	
Interest ^d	C	45	3.01	0.94	3.33	1.00	4.67	0.14	
Interest ^d	I	100	3.22	0.85	3.33	1.00	5.00	0.08	
Perceived competence ^d	P	51	2.74	0.79	2.67	1.33	4.33	0.11	-
Perceived competence ^d	A	48	2.85	0.81	3.00	1.33	5.00	0.12	
Perceived competence ^d	C	45	2.71	0.73	2.67	1.33	4.33	0.11	
Perceived competence ^d	I	100	2.70	0.70	2.67	1.00	4.00	0.07	
Perceived task importance ^d	P	51	3.72	0.92	4.00	1.00	5.00	0.13	-
Perceived task importance ^d	A	48	3.47	1.03	3.83	1.00	5.00	0.15	
Perceived task importance ^d	C	45	3.33	1.04	3.33	1.00	5.00	0.15	
Perceived task importance ^d	I	100	3.51	0.89	3.67	1.33	5.00	0.09	
Social belonging ^d	P	51	1.95	0.79	2.00	1.00	4.00	0.11	-
Social belonging ^d	A	48	2.10	0.89	2.00	1.00	4.33	0.13	
Social belonging ^d	C	45	1.94	0.80	1.67	1.00	4.00	0.12	
Social belonging ^d	I	100	2.98	0.92	3.00	1.00	5.00	0.09	
Perceived choice ^d	P	51	2.54	0.99	2.67	1.00	4.00	0.14	-
Perceived choice ^d	A	48	2.42	0.89	2.33	1.00	4.67	0.13	
Perceived choice ^d	C	45	2.27	0.93	2.33	1.00	4.33	0.14	
Perceived choice ^d	I	100	2.40	0.99	2.33	1.00	5.00	0.10	
Passive cognitive processes ^e	P	50	2.57	0.72	2.25	1.00	4.25	0.10	-
Passive cognitive processes ^e	A	48	2.15	0.77	2.12	1.00	4.50	0.11	
Passive cognitive processes ^e	C	45	2.27	0.68	2.25	1.00	5.00	0.10	
Passive cognitive processes ^e	I	100	2.40	0.79	2.25	1.00	5.00	0.08	
Active cognitive processes ^e	P	50	3.62	0.75	3.67	1.33	5.00	0.11	-
Active cognitive processes ^e	A	48	3.50	0.87	3.67	1.33	5.00	0.13	
Active cognitive processes ^e	C	45	3.77	0.62	4.00	2.33	5.00	0.09	
Active cognitive processes ^e	I	100	3.55	0.72	3.67	1.67	5.00	0.07	
Constructive cognitive processes ^e	P	50	3.39	0.90	3.67	1.33	5.00	0.13	-
Constructive cognitive processes ^e	A	48	3.30	0.85	3.50	1.33	4.67	0.12	
Constructive cognitive processes ^e	C	45	3.58	0.69	3.67	2.00	5.00	0.10	
Constructive cognitive processes ^e	I	100	3.43	0.79	3.67	1.33	5.00	0.08	
Interactive cognitive processes ^e	P	50	2.24	0.99	2.00	1.00	4.75	0.14	-
Interactive cognitive processes ^e	A	48	2.35	1.05	2.25	1.00	4.75	0.15	
Interactive cognitive processes ^e	C	45	2.53	1.02	2.75	1.00	4.75	0.15	
Interactive cognitive processes ^e	I	100	3.63	0.79	3.75	1.00	4.75	0.08	

Note.

^a 0 to 32 points.^b 0 to 34 points.^c in min.^d 1 = do not agree at all to 5 = totally agree.^e 1 = does not apply at all to 5 = fully applies.

6.4. Effect of types of learning activities on students' motivation

To explore whether students' motivation differed according to the assigned ICAP condition (RQ3), a multivariate analysis of variance (MANOVA) was conducted with learning condition (i.e., four treatment groups: passive, active, constructive, interactive) as independent variable and five self-reported motivational outcomes—interest, competence, choice, social belonging, and autonomy—as dependent variables. The multivariate effect was statistically significant, Pillai's Trace = 0.30,

$F(15, 714) = 5.25, p < .001$, indicating an overall difference between conditions.

Follow-up univariate ANOVAs revealed a statistically significant effect of condition on perceived relatedness, $F(3, 240) = 25.56, p < .001$. No statistically significant differences between conditions were observed for the other dependent variables. Post-hoc Tukey HSD tests revealed that participants in the interactive condition reported statistically significantly higher levels of perceived relatedness compared to those in the passive ($M_{\text{diff}} = 1.03, 95\% \text{ CI } [0.64, 1.41], p < .001$), active

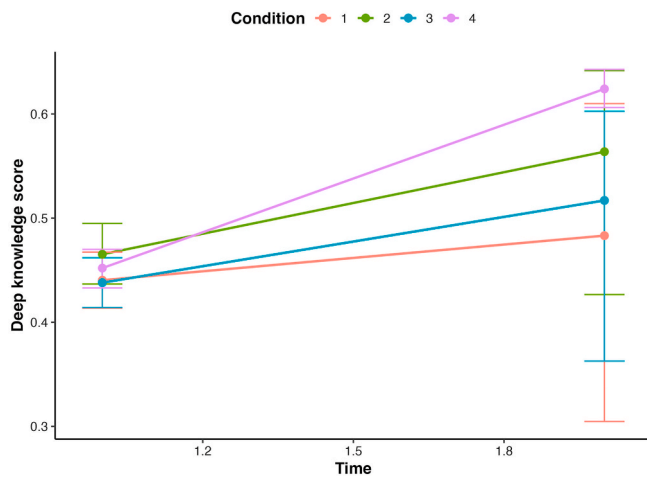


Fig. 1. Development of deep knowledge scores by condition (1 = passive, 2 = active, 3 = constructive, 4 = interactive).

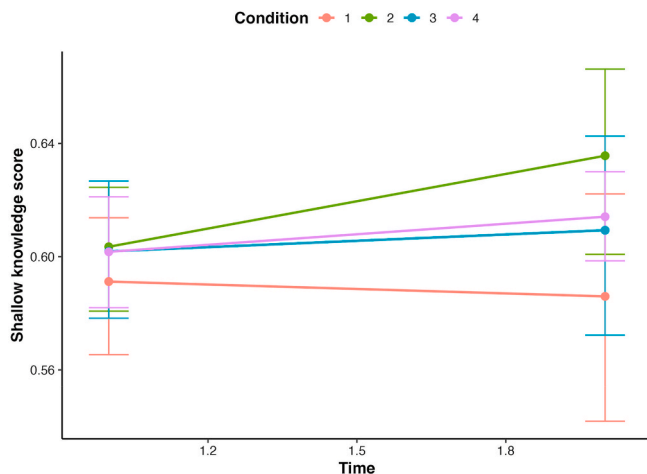


Fig. 2. Development of shallow knowledge scores by condition (1 = passive, 2 = active, 3 = constructive, 4 = interactive).

($M_{diff} = 0.88$, 95% CI [0.49, 1.27], $p < .001$), and constructive conditions ($M_{diff} = 1.04$, 95% CI [0.63, 1.44], $p < .001$). No other contrasts reached statistical significance ($ps > 0.80$). That is, the MANOVA indicated overall differences in motivation across ICAP conditions, driven solely by higher perceived relatedness in the interactive condition, whereas no other motivational outcomes differed significantly.

6.5. Effect of types of learning activities on students' cognitive processes

To explore whether students reported having engaged in the type of cognitive process aligned with their assigned ICAP condition (RQ4), a MANOVA was conducted with learning condition (i.e., four treatment groups: passive, active, constructive, interactive) as independent variable and four reported types of cognitive processes (i.e., passive, active, constructive, interactive) as dependent variables. The multivariate test was statistically significant, Pillai's Trace = 0.37, $F(15, 11) = 6.57$, $p < .001$, indicating that overall cognitive process ratings varied by assigned condition.

Follow-up univariate ANOVAs showed a statistically significant effect for reported interactive cognitive processes, $F(3, 239) = 36.72$, $p < .001$. No statistically significant effects were found for passive, active, or constructive cognitive processes, suggesting that only interactive cognitive process ratings clearly differed by assigned condition.

Post-hoc Tukey HSD tests revealed that participants in the interactive

condition reported statistically significantly higher levels of interactive cognitive processes compared to those in the passive ($M_{diff} = 1.40$, 95% CI [0.98, 1.81], $p < .001$), active ($M_{diff} = 1.28$, 95% CI [0.85, 1.70], $p < .001$), and constructive conditions ($M_{diff} = 1.10$, 95% CI [0.66, 1.53], $p < .001$). No other contrasts reached statistical significance ($ps > 0.40$). That is, the MANOVA showed that cognitive process ratings differed across conditions, driven exclusively by higher reported interactive processes in the interactive condition, with no differences for passive, active, or constructive processes.

6.6. Effect of cognitive processes on students' shallow and deep knowledge gain

To explore whether students' cognitive processes predicted learning outcomes (RQ5), multiple linear regressions were conducted for both shallow and deep knowledge gains using the four reported types of cognitive processes (passive, active, constructive, and interactive) as predictors.

For shallow knowledge gain, the overall model was not statistically significant, $F(4, 238) = 0.26$, $p = .902$, $R^2 = 0.004$. None of the ICAP predictors reached statistical significance ($ps > 0.46$), and all standardized regression coefficients were close to zero, indicating that the degree of reported types of cognitive processes did not meaningfully predict shallow knowledge gains.

For deep knowledge gain, the regression model was also not statistically significant overall, $F(4, 238) = 2.50$, $p = .098$, $R^2 = 0.04$. The predictor *constructive cognitive processes* ($\beta = -0.04$, $SE = 0.01$, $p = .010$) was negatively associated with deep knowledge gain. The other three predictors were not statistically significant.

These results proved to be robust in the second robustness check. That is, reported cognitive processes did not predict shallow knowledge gains, but showed a small, robust association with deep knowledge gains, with constructive processes negatively related to performance.

7. Discussion

In the present study, we investigated the ICAP hypothesis, along with further assumptions that evolved around the ICAP framework, to address several shortcomings of empirical ICAP-based studies to date. Overall, we found little evidence in favour of the ICAP hypothesis. The results will be discussed in more detail below.

Differences in learners' deep knowledge gain (RQ1) in the four ICAP-based conditions were statistically non-significant. Across all conditions, students showed a comparable increase in deep knowledge. In contrast, prior knowledge and time on task were revealed to be statistically significant predictors. Given that we did not observe any effects on deep knowledge gain and the ICAP framework suggesting that effects may not emerge for shallow knowledge gain, it is not surprising that no effects were found for shallow knowledge (RQ2). Thus, the ICAP hypothesis, which postulates a hierarchy of effects based on types of learning activities and their impact on knowledge gain, could not be supported.

This study is not the first to cast doubt on the validity of the ICAP hypothesis (Lam & Muldner, 2017; Luo et al., 2016; Petersen et al., 2023). One reason for the lack of effects may lie within the stricter methodological approach (particularly by controlling for time on task as an important predictor in our study) adopted in this study relative to previous studies that found support for the ICAP hypothesis. For instance, Liu et al. (2024) did not include a pretest or learning material. In addition, time-on-task varied across conditions, but was not controlled for. Whereas students in the constructive condition were instructed to construct an argument map after reading a text, students in the passive condition received only the text. They were then prompted to perform a non-learning relevant task. In contrast, other studies, such as the one by Lam and Muldner (2017), which employed a more rigorous methodology by controlling for time-on-task and incorporating both shallow and deep knowledge gain assessments, found no effects.

Accordingly, there is a need for more controlled studies exploring the ICAP hypothesis in its entirety.

The effect of types of learning activities on students' motivation (RQ3) suggests that students who were prompted to engage in interactive learning activities experienced a statistically significantly higher sense of social belonging during learning than those prompted to engage in constructive, active, and passive learning activities. In contrast, other motivational perceptions did not statistically significantly differ based on the type of learning activity. Consequently, our results seem to diverge from earlier studies (e.g., Xie et al., 2024) documenting motivational gains across the ICAP hierarchy. Yet, the results of earlier studies stem from within-subject designs, and in the study of Wekerle et al. (2024) differential effects were also only observable for one (C vs. P) learning activity comparison and one out of two motivational constructs.

Moreover, as we could only observe an effect for social belonging, we may also need to consider methodological reasons that explain the present finding. Students in the interactive condition were theoretically able to perceive this form of motivation, as they were the only ones who collaborated with their peers. Accordingly, this finding can be interpreted as a methodological artefact or evidence of a successful manipulation check. In conclusion, despite single associations, the empirical basis for aligning or extending the cognitive ICAP framework from a motivational perspective appears insufficient.

The analysis of the types of learning activities on students' cognitive processes (RQ4) suggests that participants clearly perceived and reported their engagement in the interactive condition. In contrast, the distinction between the other three types of learning activities was less evident. This suggests that, despite the prompted types of learning activities, students predominantly engaged in cognitive processes other than those theoretically expected according to the framework or in differing processes. This finding supports earlier evidence-based criticisms that challenge the idea of inferring cognitive engagement from observable behaviors (see Renkl, 2011; Thurn et al., 2023) and underscores the need for further empirical investigation of the largely unexplored relationship between observable learning activities and cognitive processes within the ICAP framework.

To further contextualize the results, analysing students' learning outputs (Chi & Boucher, 2023) or employing a think-aloud approach (Li, 2021) may prove informative. However, such methods represent only proxies for the cognitive processes that actually take place. Nonetheless, an effect on interactive cognitive processes might still have occurred, as students may have been better able to identify interactive learning based on the specific item phrasing (i.e., the reference to a learning partner), in contrast to the other three types of cognitive processes. Accordingly, the instrument of Stegmann et al. (2024) should be revised in future research to ensure that students across all conditions are equally able to report all theoretically relevant cognitive processes, thereby increasing its reliability.

Finally, regarding the effect of students' cognitive processes on their knowledge gain (RQ5), the findings indicate a slight trend that students who reported higher levels of constructive engagement showed lower gains in deep knowledge. Given that the reported cognitive processes show at least some effect on learning outcomes, compared to the types of learning activities, these processes could be considered a more proximal indicator in light of the present findings. However, the results contradict the assumptions in the ICAP framework. As there is currently a lack of empirical ICAP-based studies on students' cognitive processes, no evidence-based explanations for this finding can be provided yet. Nonetheless, instructional research repeatedly points to the importance of also considering working memory capacity or cognitive load (Sweller et al., 2019). Therefore, to better understand the contradictory findings, future studies should consider including a measure of cognitive load.

7.1. Limitations

Despite the insights provided by our study, several methodological considerations need to be taken into account when interpreting the findings:

First, similar to the initial study of Menekse et al. (2013), the present study was designed as a short-term intervention, which limits the possibility of making inferences about the effects of a more sustained engagement in the different learning activities and underlying cognitive processes. Future research might benefit from implementing more extended interventions. At the same time, such learning settings may reduce ecological validity, as learning activities rarely occur in isolation in authentic educational contexts (Wekerle et al., 2024).

Second, we used no follow-up measurement. Previous research (e.g., LaDage et al., 2018) has shown that effects may only be detectable immediately after the intervention. However, in the case of our result pattern, there is little reason to expect qualitatively different outcomes at a later measurement point.

Third, as previously noted, the reliability of the cognitive process instrument (Stegmann et al., 2024) was lower than expected. Therefore, those results should be interpreted with caution and clearly indicate the need to refine the scales.

Fourth, regarding affective-motivational outcomes, we only measured motivation in the present study and did not include affective states such as emotions. Including measures of affective states could have provided additional insights into the impact of the learning activities on students' emotional experiences; future research should address this gap.

Fifth, we did not code the open-ended responses in the learning phase provided by participants in the constructive and interactive condition which might provide insight to explain potential differences in participants' learning outcomes across conditions. This decision was made because the nature and quantity of learning outputs differed substantially across conditions: In the constructive and interactive conditions, participants generated open responses, whereas in the passive and active conditions, such outputs were either absent or not comparable in scope and content. As a result, performance could not be assessed in a standardized way across all groups. However, we recognize that future research would benefit from analysing students' text data generated during learning and from developing and validating instruments that allows researchers to compare different types of student outputs with regard to their quality of processing.

Sixth, even though we used identical pre- and posttest items to ensure direct comparability of knowledge scores across both measurement points, practice effects may have occurred and influenced potential effects.

Lastly, to ensure comparable time-on-task across conditions, students in the passive condition were instructed to reread the text. This raises the question of whether the evoked cognitive processes can still be considered passive in nature or whether the instruction to re-read the text may have prompted more elaborate cognitive processes despite the fact that the observable learning activity remained passive.

7.2. Theoretical implications

In conclusion, our findings suggest that the empirical foundation supporting the ICAP framework may not yet be sufficiently robust to recommend the ICAP framework or observable learning activities as an explanatory approach for elucidating differences in students' learning outcomes. In our study, we mostly could not observe systematic differences in either students' reported cognitive processes or their learning outcomes across conditions.

While it seems plausible that deeper learning often leads to observable active or interactive activities ($A \rightarrow B$; e.g., Chi, 2009), the reverse is less clear: engaging in an activity, such as collaboration, does not necessarily produce deeper learning ($B \nrightarrow A$; e.g., Renkl, 2011). It

depends on the quality of the collaboration, whether it is effectively leading to deeper learning. Technology may strengthen this link by shaping the quality of interaction, making deeper learning more likely (Sailer et al., 2024). For instance, Fütterer et al. (2023) found that when students engaged in learning activities using digital tools, these activities were associated with higher learning effort, suggesting that the ICAP hypothesis might be more accurate in technology-enhanced teaching situations.

7.3. Methodological implications

Methodologically, the study demonstrates the importance of experimental between-subjects designs that include all four types of learning activities, use identical learning materials, control for task duration, and employ both shallow-level and deep-level knowledge assessments. By also examining students' cognitive processes and associations with both the type of learning activity and the resulting outcomes, this study marks an important first step for a comprehensive evaluation of the ICAP framework. Yet, more research is needed that applies a more rigorous methodology to explore the ICAP framework and includes a broader range of measures – such as triangulated assessments of cognitive processes and cognitive load – to capture the complexity of students' learning activities and their effects more effectively. While such controlled designs present clear strengths, it would be valuable to conduct studies in more authentic learning contexts as well that also use more extended time frames. This might be of particular interest in the domain of self-regulated learning (e.g., in preparation for exams), which becomes increasingly central from primary and secondary to higher education.

7.4. Practical implications

Based on our findings, we are currently unable to offer robust recommendations for instructional design and educational practice. At this stage, our results may primarily serve to remind educators that simply implementing interactive or constructive learning activities does not automatically lead to deeper learning; the quality and context of these

activities might be crucial. Therefore, the results encourage practitioners to adopt a nuanced approach, ensuring that instructional strategies are aligned with both the content and the learners' needs.

CRedit authorship contribution statement

Christina Wekerle: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Michael Sailer:** Writing – review & editing, Methodology, Conceptualization. **Ingo Kollar:** Writing – review & editing, Conceptualization. **Tim Fütterer:** Writing – review & editing, Visualization, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly (v1.114.5.0) and ChatGPT-4o (v1.2025.098) to ensure the absence of spelling errors and to obtain suggestions for more suitable phrasing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research was funded by Stiftung Innovation in der Hochschullehre (Grant FBM2020-EA-2620-01350) and was supported by the Postdoctoral Academy of Education Sciences and Psychology of the Hector Research Institute of Education Sciences and Psychology, Tübingen, funded by the Baden-Württemberg Ministry of Science, Research, and the Arts.

Acknowledgement

We would like to thank Katharina Liebl for her support in preparing the materials and implementing the study.

Appendix

Link to the OSF project: <https://doi.org/10.17605/OSF.IO/KJU36>.

Correlation tables for each condition

Correlation matrix for the passive condition

Variable	Deep knowledge T1	Deep knowledge T2	Surface knowledge T1	Surface knowledge T2	Time on task (in sec)	Interest	Perceived competence	Perceived task importance	Social belonging	Perceived choice	Elaboration strategy	Passive cognitive processes	Active cognitive processes	Constructive cognitive processes	Interactive cognitive processes
Deep knowledge T1	1.00	0.40	−0.33	0.12	−0.31	0.27	0.25	0.32	0.12	0.37	0.17	−0.01	0.19	0.23	0.19
Deep knowledge T2		1.00	0.01	0.81	−0.15	0.28	0.14	0.33	0.10	−0.07	0.32	−0.09	0.21	0.04	−0.22
Shallow knowledge T1			1.00	0.24	−0.20	0.06	0.09	0.07	−0.11	−0.36	0.08	−0.37	0.13	0.16	−0.05
Shallow knowledge T2				1.00	0.07	0.22	0.14	0.23	−0.14	−0.14	0.28	−0.21	0.13	0.17	−0.20
Time on task (in sec)					1.00	−0.21	−0.17	−0.30	0.14	0.08	−0.32	0.20	−0.11	−0.20	−0.07
Interest						1.00	0.50	0.80	0.42	0.22	0.47	−0.47	0.55	0.57	0.01
Perceived competence							1.00	0.40	0.41	0.32	0.45	−0.47	0.49	0.47	0.15
Perceived task importance								1.00	0.45	0.17	0.46	−0.39	0.33	0.49	0.21
Social belonging									1.00	0.43	0.16	−0.15	0.12	0.16	0.41
Perceived choice										1.00	0.05	−0.03	0.28	0.29	0.30
Elaboration strategy											1.00	−0.48	0.48	0.62	0.12
Passive cognitive processes												1.00	−0.24	−0.28	0.08
Active cognitive processes													1.00	0.66	0.07

(continued on next page)

(continued)

Variable	Deep knowledge T1	Deep knowledge T2	Surface knowledge T1	Surface knowledge T2	Time on task (in sec)	Interest	Perceived competence	Perceived task importance	Social belonging	Perceived choice	Elaboration strategy	Passive cognitive processes	Active cognitive processes	Constructive cognitive processes	Interactive cognitive processes
Constructive cognitive processes														1.00	0.18
Interactive cognitive processes															1.00

Note: Pearson correlations (pairwise complete observations). Upper-triangular matrix shown; diagonal = 1.00.

Correlation matrix for the active condition

Variable	Deep knowledge T1	Deep knowledge T2	Surface knowledge T1	Surface knowledge T2	Time on task (in sec)	Interest	Perceived competence	Perceived task importance	Social belonging	Perceived choice	Elaboration strategy	Passive cognitive processes	Active cognitive processes	Constructive cognitive processes	Interactive cognitive processes
Deep knowledge T1	1.00	0.36	-0.03	0.20	0.07	0.12	0.01	0.09	-0.08	0.17	0.01	0.10	-0.09	-0.08	-0.02
Deep knowledge T2		1.00	0.21	0.75	-0.07	0.03	-0.09	0.04	-0.05	-0.02	0.03	-0.01	0.12	0.02	0.21
Shallow knowledge T1			1.00	0.24	-0.19	0.05	0.12	0.08	-0.15	0.06	-0.08	-0.18	0.05	-0.12	-0.13
Shallow knowledge T2				1.00	-0.23	0.08	0.22	0.02	-0.14	-0.07	0.07	-0.20	0.03	-0.12	-0.15
Time on task (in sec)					1.00	0.06	-0.16	0.10	0.07	-0.00	0.08	-0.00	0.11	0.01	0.07
Interest						1.00	0.48	0.76	0.07	0.35	0.65	-0.49	0.41	0.44	0.04
Perceived competence							1.00	0.45	0.07	0.21	0.21	-0.30	0.14	0.24	-0.03
Perceived task importance								1.00	-0.32	0.00	0.61	-0.42	0.35	0.29	0.01
Social belonging									1.00	0.39	-0.09	0.28	-0.09	0.10	0.29
Perceived choice										1.00	-0.09	-0.04	0.16	0.13	0.13
Elaboration strategy											1.00	-0.44	0.53	0.61	0.09
Passive cognitive processes												1.00	-0.25	-0.36	0.21
Active cognitive processes													1.00	0.68	0.11
Constructive cognitive processes														1.00	0.19
Interactive cognitive processes															1.00

Note: Pearson correlations (pairwise complete observations). Upper-triangular matrix shown; diagonal = 1.00.

Correlation matrix for the constructive condition

Variable	Deep knowledge T1	Deep knowledge T2	Surface knowledge T1	Surface knowledge T2	Time on task (in sec)	Interest	Perceived competence	Perceived task importance	Social belonging	Perceived choice	Elaboration strategy	Passive cognitive processes	Active cognitive processes	Constructive cognitive processes	Interactive cognitive processes
Deep knowledge T1	1.00	0.37	-0.13	0.27	0.02	0.04	0.23	0.08	-0.12	0.07	-0.00	0.04	-0.17	-0.23	-0.00
Deep knowledge T2		1.00	0.29	0.77	0.12	0.00	0.15	0.23	-0.12	-0.12	-0.04	-0.29	-0.12	-0.03	-0.15
Shallow knowledge T1			1.00	0.41	-0.09	0.21	-0.14	0.20	0.05	-0.26	0.20	-0.40	0.18	0.18	-0.16
Shallow knowledge T2				1.00	-0.04	0.09	-0.28	0.13	-0.12	-0.30	0.02	-0.16	0.18	0.05	0.05
Time on task (in sec)					1.00	0.22	0.13	0.28	-0.00	0.34	0.10	-0.01	0.32	0.28	0.08
Interest						1.00	0.30	0.81	0.21	0.24	0.35	-0.37	0.53	0.72	-0.15
Perceived competence							1.00	0.25	0.05	0.40	0.23	-0.12	0.13	0.05	0.02
Perceived task importance								1.00	0.10	0.32	0.36	-0.51	0.40	0.61	-0.26
Social belonging									1.00	0.07	-0.16	-0.06	0.16	0.12	0.00
Perceived choice										1.00	0.08	0.27	0.12	0.11	0.15
Elaboration strategy											1.00	-0.39	0.58	0.53	0.03
Passive cognitive processes												1.00	-0.30	-0.51	0.09
Active cognitive processes													1.00	0.68	0.21
Constructive cognitive processes														1.00	0.03
Interactive cognitive processes															1.00

Note: Pearson correlations (pairwise complete observations). Upper-triangular matrix shown; diagonal = 1.00.

Correlation matrix for the interactive condition

Variable	Deep knowledge T1	Deep knowledge T2	Surface knowledge T1	Surface knowledge T2	Time on task (in sec)	Interest	Perceived competence	Perceived task importance	Social belonging	Perceived choice	Elaboration strategy	Passive cognitive processes	Active cognitive processes	Constructive cognitive processes	Interactive cognitive processes
Deep knowledge T1	1.00	0.18	−0.12	−0.07	−0.05	0.05	0.07	−0.01	−0.18	0.02	0.18	−0.06	0.17	0.34	0.02
Deep knowledge T2		1.00	0.01	0.16	0.03	0.10	0.12	0.04	−0.08	0.04	0.23	−0.06	0.15	0.08	0.08
Shallow knowledge T1			1.00	0.39	−0.01	0.05	−0.06	0.04	0.03	0.08	0.08	0.20	0.04	−0.10	−0.03
Shallow knowledge T2				1.00	0.21	0.22	0.05	0.09	−0.03	0.02	0.25	−0.04	0.02	0.00	0.02
Time on task (in sec)					1.00	0.27	0.22	0.04	0.05	0.04	0.06	−0.26	0.05	0.02	0.00
Interest						1.00	0.37	0.73	0.38	0.21	0.56	−0.51	0.39	0.39	0.43
Perceived competence							1.00	0.37	0.18	0.20	0.06	−0.33	−0.14	0.10	0.16
Perceived task importance								1.00	0.45	0.20	0.48	−0.44	0.39	0.44	0.41
Social belonging									1.00	0.22	0.24	−0.25	0.12	0.23	0.23
Perceived choice										1.00	−0.00	−0.01	−0.11	−0.03	0.07
Elaboration strategy											1.00	−0.37	0.67	0.64	0.43
Passive cognitive processes												1.00	−0.30	−0.46	−0.27
Active cognitive processes													1.00	0.61	0.34
Constructive cognitive processes														1.00	0.40
Interactive cognitive processes															1.00

Note: Pearson correlations (pairwise complete observations). Upper-triangular matrix shown; diagonal = 1.00.

References

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37 (45). <https://doi.org/10.1007/s10648-025-10020-8>

Chi, Micheline T. H., Adams, Joshua, Bogusch, Emily B., Bruchok, Christiana, Kang, Seokmin, Lancaster, Matthew, Levy, Roy, Li, Na, McEllood, Katherine L., Stump, Glenda S., Wylie, Ruth, Xu, Dongchen, & Yaghmourian, David L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, 42(6), 1777–1832. <https://doi.org/10.1111/cogs.12626>. <https://onlinelibrary.wiley.com/doi/10.1111/cogs.12626>. (Accessed 28 June 2018)

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>

Chi, M. T. H., & Boucher, N. S. (2023). Applying the ICAP framework to improve classroom learning. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 94–110). Society for the Teaching of Psychology.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>

Conley, Q., Sadauskas, J., Christopherson, R., Lin, L., Ilgaz, H., Seto, C., Kula, I., Dalal, M., & Atkinson, R. K. (2022). Facebook usage patterns looking into the mind via the ICAP engagement framework. *Behaviour & Information Technology*, 42(5), 514–526. <https://doi.org/10.1080/0144929X.2021.2024597>

Craik, Fergus I. M., & Tulving, Endel (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>. <https://doi.org/10.1037/0096-3445.104.3.268>. (Accessed September 1975)

Craik, Fergus I. M., & Lockhart, Robert S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/s0022-5371\(72\)80001-x](https://doi.org/10.1016/s0022-5371(72)80001-x). <https://linkinghub.elsevier.com/retrieve/pii/S002253717280001X>

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer. <https://doi.org/10.1007/978-1-4899-2271-7>

Fütterer, T., Hoch, E., Lachner, A., Scheiter, K., & Stürmer, K. (2023). High-quality digital distance teaching during COVID-19 school closures: Does familiarity with technology matter? *Computers & Education*, 199, Article 104788. <https://doi.org/10.1016/j.compedu.2023.104788>

Greene, Barbara A. (2015). Measuring Cognitive Engagement With Self-Report Scales: Reflections From Over 20 Years of Research. *Educational Psychologist*, 50(1), 14–30. <https://doi.org/10.1080/00461520.2014.989230>. <http://www.tandfonline.com/doi/abs/10.1080/00461520.2014.989230>. (Accessed 13 January 2015)

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. <https://www.jstor.org/stable/4615733>.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

LaDage, L. D., et al. (2018). Variation in behavioral engagement during an active learning activity leads to differential knowledge gains in college students. *Advances in Physiology Education*, 42(1), 99–103. <https://doi.org/10.1152/advan.00150.2017>

Lam, R., & Muldner, K. (2017). Manipulating cognitive engagement in preparation-to-collaborate tasks and the effects on learning. *Learning and Instruction*, 52, 90–101. <https://doi.org/10.1016/j.learninstruc.2017.05.002>

Lenth, R. V. (2024). Emmeans: Estimated marginal means, aka least-squares means. *Version R package version 1.10.0* [Computer software]. <https://CRAN.R-project.org/package=emmeans>.

Leutner, D., Opfermann, M., & Schmeck, A. (2014). Lernen mit Medien. In T. Seidel, & A. Krapp (Eds.), *Pädagogische psychologie* (6th ed., pp. 297–322). Beltz.

Li, S. (2021). Measuring cognitive engagement: An overview of measurement instruments and techniques. *International Journal of Psychology and Educational Studies*, 8(3), 63–76. <https://doi.org/10.52380/ijpes.2021.8.3.239>

Liu, Q., Zhong, Z., & Nesbit, J. C. (2024). Argument mapping as a pre-writing activity: Does it promote writing skills of EFL learners? *Education and Information Technologies*, 29, 7895–7925. <https://doi.org/10.1007/s10639-023-12098-5>

Luo, L., Kiewra, K. A., & Samuelson, L. (2016). Revising lecture notes: How revision, pauses, and partners affect note taking and achievement. *Instructional Science*, 44, 45–67. <https://doi.org/10.1007/s11251-016-9370-4>

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, 102(3), 346–374. <https://doi.org/10.1002/jee.20021>

Petersen, G. B., Stenberdt, V., Mayer, R. E., & Makransky, G. (2023). Collaborative generative learning activities in immersive virtual reality increase learning. *Computers & Education*, 207(104931). <https://doi.org/10.1016/j.compedu.2023.104931>

Piedmont, R. L. (2023). Inter-item correlations. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3577–3578). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_1493

Renkl, A. (2011). Aktives Lernen: Von sinnvollen und weniger sinnvollen theoretischen Perspektiven zu einem schillernden Konstrukt [Active learning: From sensible and less sensible theoretical perspectives to an ambiguous construct]. *Unterrichtswissenschaft*, 39(3), 197–212.

Sailer, M., Henze, J. U., Mayer, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69, 371–380. <https://doi.org/10.1016/j.chb.2016.12.033>

Sailer, M., Maier, R., Berger, S., Kastorff, T., & Stegmann, K. (2024). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences*, 112, Article 102446. <https://doi.org/10.1016/j.lindif.2024.102446>

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, 60. <https://doi.org/10.1016/j.newideapsych.2020.100832>

Stegmann, K., Berger, S., Bannert, M., Fischer, F., Gräsel, C., Lindner, M., Nerdel, C., Neuhaus, B., Haldenwang, V., & Ufer, S. (2024). *DigitUS. Digitalisierung von Unterricht in der Schule. Abschlussbericht*. Ludwig-Maximilians-Universität München.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. <https://doi.org/10.1023/A:1022193728205>
- Thurn, C. M., Edelsbrunner, P. A., Berkowitz, M., Deiglmayr, A., & Schalk, L. (2023). Questioning central assumptions of the ICAP framework. *Science of Learning*, 8(49). <https://doi.org/10.1038/s41539-023-00197-4>
- Wekerle, C., Daumiller, M., Janke, S., Dickhäuser, O., Dresel, M., & Kollar, I. (2024). Putting ICAP to the test: How technology-enhanced learning activities are related to cognitive and affective-motivational learning outcomes in higher education. *Scientific Reports*, 14(16295). <https://doi.org/10.1038/s41598-024-66069-y>
- What Works Clearinghouse. (2022). What works clearinghouse procedures and standards handbook. U.S. department of education, institute of education sciences, national center for education evaluation and regional assistance (NCEE) version 5.0. https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 31–45. <https://doi.org/10.25656/01:31663>
- Xie, H., Lin, D., He, W., & Chen, Q. (2024). The aesthetics at a pencil tip: The effects of drawing on learning poems. *Learning and Instruction*, 91(101881). <https://doi.org/10.1016/j.learninstruc.2024.101881>
- Christina Wekerle was a research associate at the University of Augsburg. Her research interests entail technology-enhanced learning and teaching, instructional design, and scientific reasoning.
- Michael Sailer is full professor at the University of Augsburg. His research interests entail technology-enhanced learning and teaching, gamified and simulation-based learning, and AI in education.
- Tim Fütterer is a postdoctoral researcher at the University of Tübingen. His research interests entail technology-enhanced teaching, teacher professional development, and AI in education.